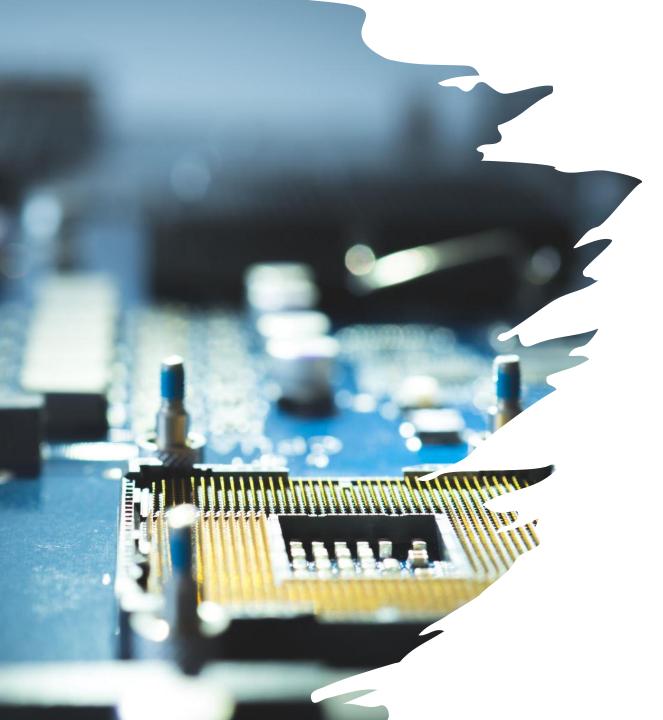




Sistemas Operacionais

Hardware

MEMÓRIAS (continuação)



- Historicamente as CPU sempre foram mais rápidas que memórias;
 - Memórias melhoraram e CPU também;
 - Desequilíbrio permanece.

Projetistas costumam usar novas tecnologias para aumentar a capacidade das memórias e não a velocidade.

CPU faz uma requisição e aguarda vários ciclos de CPU para ter resposta da memória.

Problema econômico e não tecnológico.

- Há tecnologia para se construir memórias tão rápidas quanto o processador mas elas tem que ficar localizadas dentro do processador.
 - Dados trafegados em barramento geram uma operação mais lenta.

Instalar memória em um chip de CPU faz com que ela fique maior;

• Custo aumenta;

Opção:

- Pequena quantidade de memória rápida ou
- Grande quantidade de memória lenta

Solução intermediária:

- Memórias CACHE
- Conceitos importantes:
 - Latência: tempo de demora na entrega de um operando.
 - Largura de banda: quantidade de dados fornecida por unidade de tempo.

Memória CACHE:

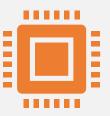
- Palavras de memória usados com mais frequência são mantidos em uma memória intermediária chamada: cache.
- Quando a CPU precisa de uma palavra ela examina primeiro a cache.

Obs.: **Palavra** é a unidade natural de informação usada por cada tipo de computador.

É uma sequência de bits de tamanho fixo

A quantidade de dados transferidos entre os processadores e a memória é também geralmente uma palavra

Memórias CACHE -Resumo



Dispositivo de acesso rápido, interno a um sistema, que serve de intermediário entre um operador de um processo e o dispositivo de armazenamento ao qual esse operador acede.



Vantagem: Evitar acesso ao dispositivo de armazenamento - que pode ser lento

Acesso dos programas na memória:

- Programas não acessam suas memórias em total aleatoriedade.
- Se uma dada referência à memória for para o endereço A, é muito provável que a próxima referência à memória estará na vizinhança geral de A.

Princípio da localidade:

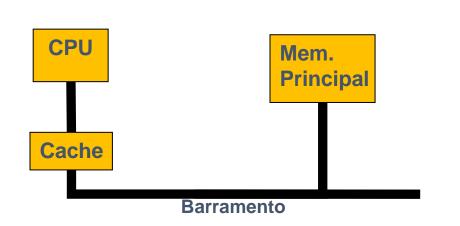
- Um programa acessa uma porção relativamente pequena do espaço endereçável em um instante qualquer.

Localidade temporal: Se um item é referenciado, ele tenderá a ser referenciado novamente.

Localidade espacial: Se um item é referenciado, itens cujos endereços são próximos a este, tenderão a ser referenciados também.

Ideia geral: quando uma palavra for referenciada, ela e algumas de suas vizinhas sejam trazidas da memória grande e lenta para a cache.

Ideia geral: quando uma palavra for referenciada, ela e algumas de suas vizinhas são trazidas da memória grande e lenta para a cache, <u>de modo que, da próxima vez que for usada, será acessada rapidamente.</u>



Ex.: caixas d'água.

L1: banheiro

L2: casa

L3: bairro

Principal: cidade (barragem)

Arranjo: CPU + Cache + Memória Principal

Conceitos embutidos:

cache miss: sinalizar que o dado não foi encontrado e seguir o processo de busca na memória principal;

cache hit: sinalizar que o dado foi encontrado na cache e bloquear a busca na memória principal.

Alternativa em alguns sistemas:

Referenciar a cache e memória principal ao mesmo tempo para ganhar tempo caso a cache não tenha a palavra desejada.

Problema:

- A implantação deste modelo é mais complicada em termos de arquitetura.

Princípio da localidade (outras considerações):

- -Memórias são divididas em blocos de tamanho fixo;
- -Blocos dentro da cache são chamados de linha de cache;
- -SE (busca na cache falhar) ENTÃO
 - -TODA a linha de cache é carregada da memória principal para a memória cache e não apenas a palavra que se quer.

-Ex.:

- -Linha de cache de 64 bytes.
- -Referência ao endereço 260.
- -Qual bloco de endereço será carregado para a cache?-256 ao 319

Escala de endereços. 260 está neste bloco.

0 - 63 64 - 127 128 - 191 192 - 255 256 - 319 ←

Projeto de CACHE:

Tamanho total: quanto maior a cache melhor seu funcionamento e maior o custo;

Tamanho da linha de cache: uma cache de 16KB pode ser dividida em 1.024 linhas de 16 bytes ou 2.048 linhas de 8 bytes e etc.

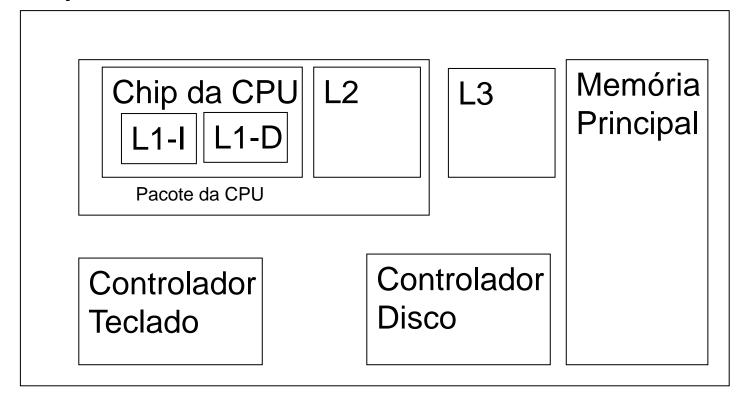
• Cache = (nº linhas * tam. bloco)

Organização: como ela controla quais palavras estão sendo mantidas no momento (Mapeamento Direto, LRU – Least Recently Used; Menos Usada Recentemente);

Instruções x Dados: usar uma <u>cache unificada</u> para ambos itens ou uma <u>cache</u> <u>dividida</u> para instruções e outra para dados.

Quantidade de cache: primária (dentro do chip), secundária (dentro do pacote do chip), terciária (externa ao chip).

Projeto de CACHE:



L1 – cache dividida entre instrução e dados.

L2 – cache unificada dentro do pacote da CPU.

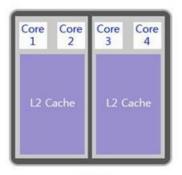
L3 – cache unificada dentro da placa mãe.

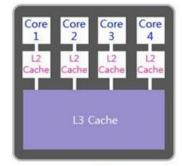
Projeto de CACHE:

- -Todas as caches são inclusivas
 - -O conteúdo de L1 está em L2
 - -O conteúdo de L2 está em L3

-...

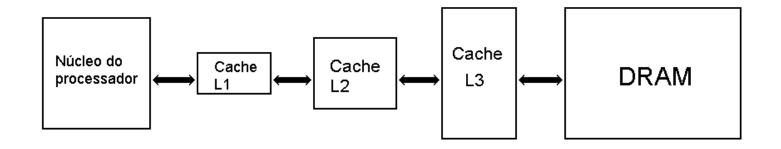


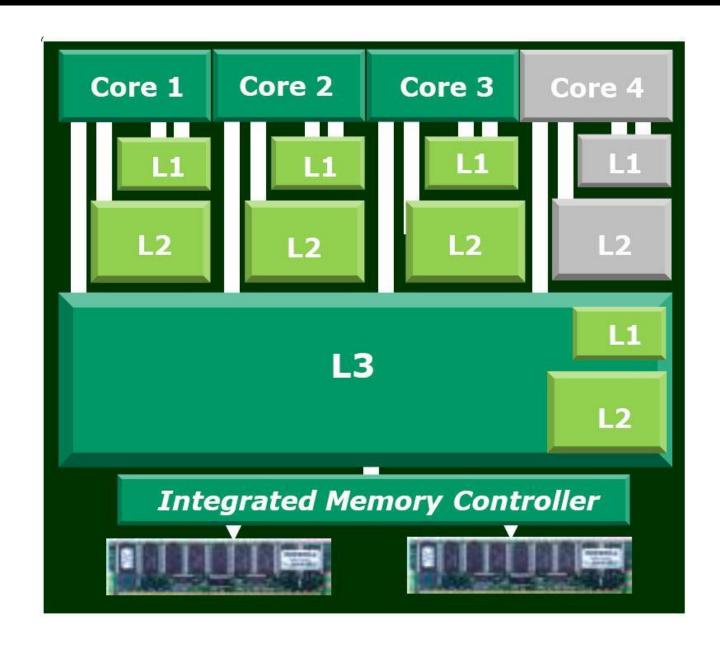












L1, L2 e L3

- Cache dividido em níveis
- A construção de memórias caches é um processo extremamente complexo e de alto custo.
- <u>Solução</u>: Construção das memórias em níveis que se diferem na relação tamanho X desempenho.

L1: Uma pequena porção de memória estática presente dentro do processador. Em alguns tipos de processador, o L1 é dividido dois níveis: Dados e Instruções (que "dizem" o que fazer com os dados).

Curiosidade: O primeiro processador da Intel a ter o cache L1 foi o i486 com 8KB.

L2: Possuindo o Cache L1 um tamanho reduzido e não apresentando uma solução ideal, foi desenvolvido o cache L2, que contém mais memória que o cache L1.

Ela é mais um caminho para que a informação requisitada não tenha que ser procurada na lenta memória principal.

Alguns processadores colocam esse cache fora do processador, por questões econômicas, pois um cache grande implica num custo grande.

L3: Terceiro nível de cache de memória.

Inicialmente utilizado pelo AMD K6-III (por apresentar o cache L2 integrado ao seu núcleo) utilizava o cache externo presente na *motherboard* como uma memória de cache adicional.

AMD Ryzen™ Threadripper™ 3990X

N° de núcleos de CPU: 64

N° de threads: 128

Clock básico: 2.9GHz

Clock de Max Boost 1: Até 4.3GHz

Cachê L1 total: 4MB

Cachê L2 total: 32MB

Cachê L3 total: 256MB

Desbloqueado 1: Sim

CMOS: TSMC 7nm FinFET

Package: sTRX4

Versão do PCI Express: PCIe 4.0

Solução térmica (PIB): Not

included

TDP / TDP Padrão: 280W

Temps máx: 95°C

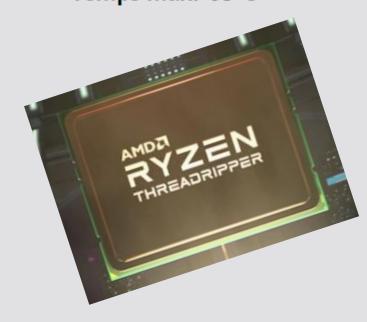
*Suporte de SO

Edição Windows 10 - 64-Bit

RHEL x86 64-Bit

*O suporte ao sistema operacional (SO) poderá variar de acordo com o

fabricante.



AMD EPYCTM 7763

General Specifications

Platform: Server

of CPU Cores: 64

Base Clock: 2.45GHz

AMD Configurable TDP (cTDP): 225-280W

Launch Date: 3/15/2021

Product Family: AMD EPYC™

of Threads: 128

L3 Cache: 256MB

CPU Socket: SP3

Product Line: AMD EPYC™ 7003 Series

Max. Boost Clock: Up to 3.5GHz

Default TDP: 280W

Socket Count: 1P/2P

Connectivity

PCI Express® Version: PCIe 4.0 x128

System Memory Specification:

Up to 3200MHz

System Memory Type: DDR4

Per Socket Mem BW: 204.8 GB/s

Memory Channels: 8



AMD EPYC™ 9754

AMD EPYC™ 9754

General Specifications

Regional Availability: Global

Plataforma: Servidor

Família de produto: AMD EPYC™

Linha de produto: AMD EPYC™ 9004 Series

Nº de núcleos de CPU: 128

N° de threads: 256

Clock de Max Boost : Até 3.1GHz

Velocidade de inicialização de todos os nú-

Clock básico: 2.25GHz

cleos 1: 3.1 GHz

Cachê L3 total: 256MB

1kU Pricing: 11,900 USD

TDP / TDP Padrão: 360W

AMD Configurable TDP (cTDP): 320-400W

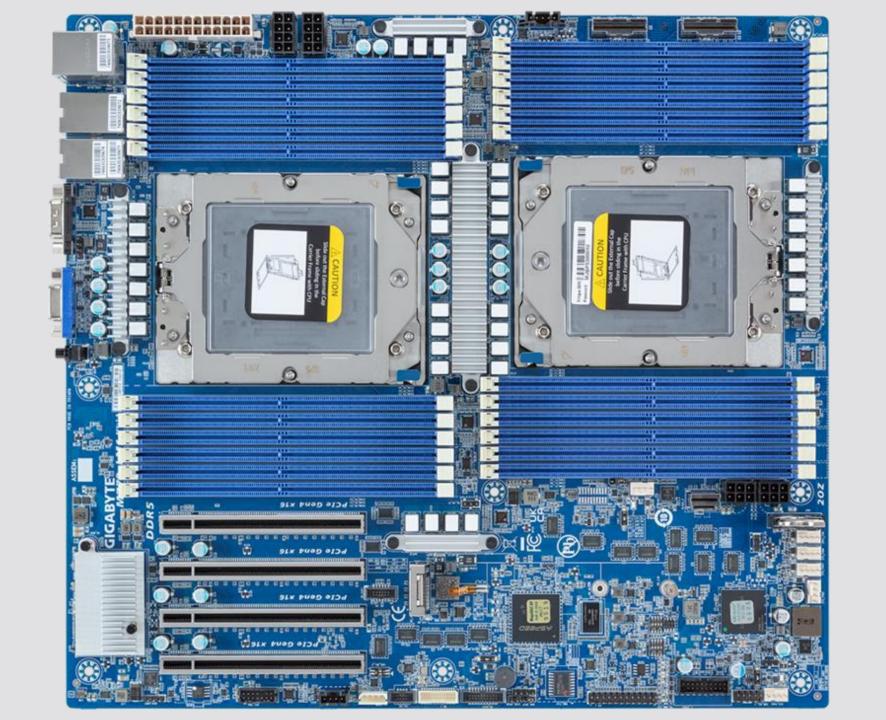
CPU Socket: SP5

Contagem de soquetes: 1P / 2P

Launch Date: 13/6/2023











Processador Intel Xeon Platinum 8280L de 28 núcleos de, 2.7GHz 28C/56T, 10.4GT/s, 38.5M Cache, Turbo, HT (205W) DDR4-2933

O processador Intel Xeon® Platinum 8280L de 28 núcleos de 2.7GHz melhora a performance e a velocidade do seu sistema. Além disso, a tecnologia de virtualização permite a migração de mais ambientes. Oferece suporte à tecnologia SpeedStep® avançada, ... Mostrar mais

R\$223.159,00

Entrega estimada para: **segunda-feira, 25 de julho** com **Envio padrão**Data estimada de entrega para 96075810

Formas de pagamento

Em até 10x sem juros de R\$ 22.315,90. Valor total a prazo R\$ 223.159,00.

1 🗸

Adicionar ao carrinho



Processador Intel Xeon Platinum 8380 de 40 núcleos de, 2.30GHz 40C/80T, 11.2GT/s, 60M Cache, Turbo, HT (270W) DDR4-3200

O processador Intel Xeon® Platinum 8380 de 40 núcleos de 2.30GHz melhora a performance e a velocidade do seu sistema. Além disso, a tecnologia de virtualização permite a migração de mais ambientes. Oferece suporte à tecnologia SpeedStep® avançada, ... Mostrar mais

R\$121.188,00

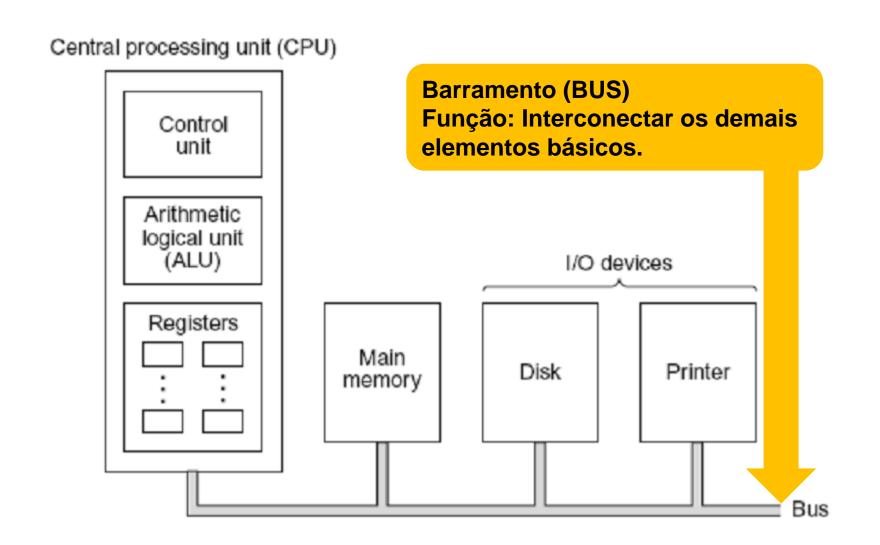
Entrega estimada para: **sexta-feira, 3 de junho** com **Envio padrão** Data estimada de entrega para 96075810

Formas de pagamento

Em até 10x sem juros de R\$ 12.118,80. Valor total a prazo R\$ 121.188,00.

1 🔻

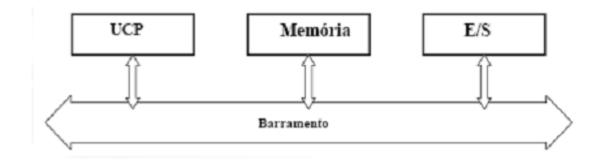
Adicionar ao carrinho



Nesse modelo, é possível transferência direta de dados entre os dispositivos de E/S e a memória

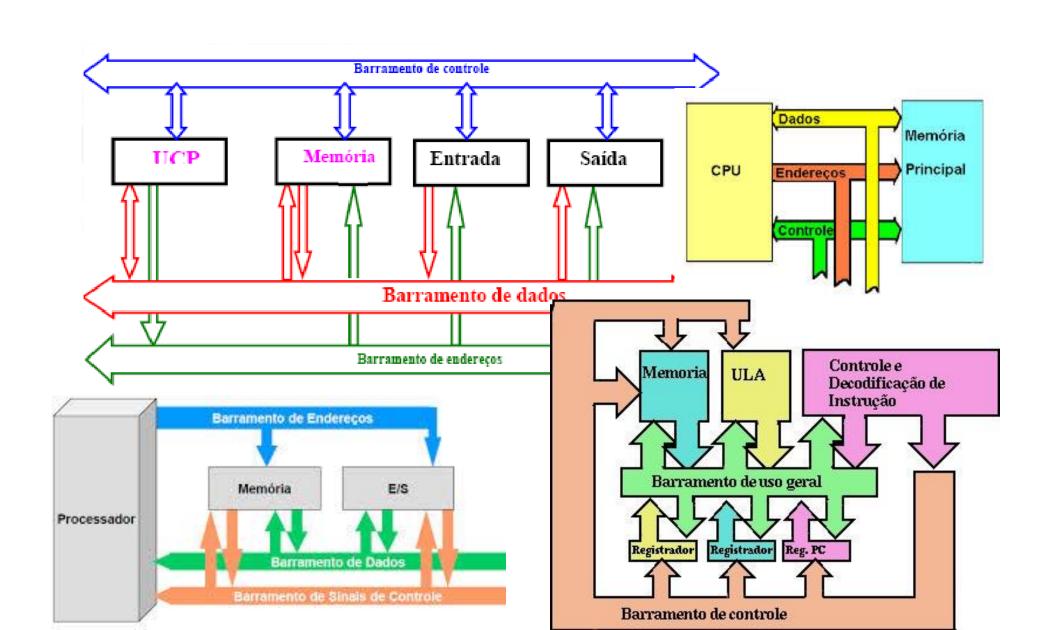
Transferência feita por dispositivos especiais

- Controlador de DMA (Direct Memory Access) ou processador E/S
- Assumem momentaneamente o controle do barramento e "isolam" a comunicação CPU com a memória e dispositivos E/S.



- Unidirecionais: conduzir os sinais em uma única direção
- Bidirecionais: conduzir ora em uma direção e ora em outra de acordo com comandos eletrônicos
- Podem ainda ser:
 - Três estados: bidirecionais acrescidos da possibilidade de se desconectarem eletricamente
 - Unidirecional com a possibilidade de desconexão elétrica
 - A possibilidade de desconexão elétrica controlada eletronicamente torna viável a comunicação seletiva entre os dispositivos.
 - Por exemplo, a CPU se comunica com a memória e se desconecta dos dispositivos E/S.

- -Tipos de barramentos:
 - Barramento de Dados: responsável pela transferência de dados e instrução entre os dispositivos.
 - É de três estados entre a memória e a CPU e unidirecional com a possibilidade de desconexão elétrica nos dispositivos E/S.
 - Barramento de Endereços: conduz o endereço a ser selecionado na memória ou dispositivos E/S.
 - É geralmente unidirecional entre a CPU (que define o endereço) e a memória e os dispositivos E/S.
 - Barramento de Controle: transfere os sinais de controle que ativam ou desativam os dispositivos, que selecionam determinado modo de operação ou sincronizam os circuitos.
- Formam a Arquitetura de Barramentos

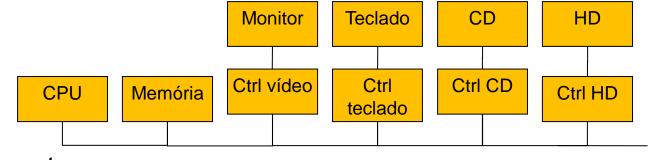


Fisicamente: todos os computadores tem estrutura similar.

Grande placa de circuito impresso na parte inferior: placa-mãe

- -CHIP CPU
- -Módulos de memória
- -Conectores para as placas de E/S

Barramento: trilha de circuito impresso que interliga os elementos acima citados.



Barramento

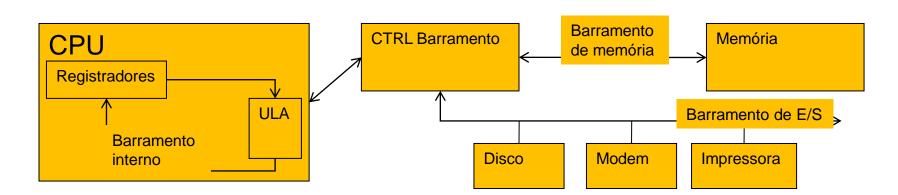
Estrutura lógica de um computador pessoal simples.

Outra definição: caminho elétrico comum entre vários dispositivos.

Podem ser internos (para transportar dados de / para ULA) ou externos (interligando dispositivos de E/S à CPU);

Inicialmente os computadores tinham somente um barramento externo, ou **barramento de sistema**, que consistia entre 50 e 100 fios de cobres gravados na placa mãe para interligar: CPU, memória e dispositivos de E/S.

Computadores modernos tem mais de um barramento externo. <u>Um de uso especial</u> entre CPU e memória e outro para dispositivos de E/S.



Protocolo do barramento: são as regras de funcionamento do barramento. Todos os dispositivos ligados ao barramento tem que respeitar essas regras (protocolo do barramento).

Dispositivos de entrada e saída:

Controlador: é uma placa de componentes eletrônicos que faz a interface entre o dispositivo e a CPU.

- -On-board: dispositivos integrados na placa-mãe do computador
- -Off-board: dispositivos avulsos vendidos separadamente da placa-mãe

Controladores: sua função é controlar dispositivos de E/S e manipular o acesso ao barramento.

Quando um controlador lê ou escreve dados de/ou para a memória sem intervenção da CPU, diz-se que ele está executando um **acesso direto à memória** (*Direct Memory Access* – **DMA**).

Concluído o uso da memória, o controlador gera uma interrupção (sinal de alerta para a CPU).

Ligação: CPU x memória x barramento:

S.O. solicita um dado que não está presente na memória principal

Controlador de disco é acionado

HD busca o dado e repassa para o controlador

Controlador faz um DMA e coloca o dado na memória

Controlador gera uma interrupção

CPU executa a rotina de interrupção

Verifica erros, executa qualquer ação especial necessária e informa o S.O. que a E/S está concluída CPU continua com o programa que foi suspenso quando ocorreu a interrupção

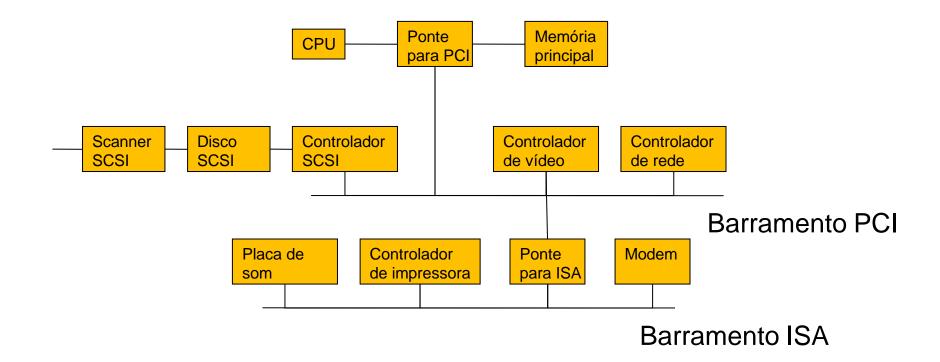
Conflito: CPU x controladores de E/S

O que acontece se CPU e controladores de E/S resolvem usar o barramento ao mesmo tempo? Um chip "árbitro de barramento" decide o que acontece.

Árbitro de barramento: chip eletrônico que define a prioridade quando CPU e controladores de E/S entram em conflito no uso do barramento.

Em geral, dispositivos de E/S tem preferência sobre CPU pois discos e dispositivos em movimento não podem ser interrompidos sob risco de danificá-los ou haver perda de dados.

<u>Importante</u>: sempre que a CPU estiver executando e usando o barramento e os dispositivos de E/S solicitarem o uso deste, eles terão prioridade e farão uso do barramento. Este processo é chamado "**roubo de ciclo**" e reduz a velocidade do computador.



PC com barramentos híbridos.

- Ativos x Passivos:
- Alguns dispositivos ligados ao barramento são <u>ativos</u> <u>e podem iniciar uma transferência</u>.
- Outros são <u>passivos</u> e apenas <u>aguardam requisições</u> <u>vindas do barramento</u>.
- Ex.: CPU ordena a gravação de um dado no disco.
 - CPU (ativo)
 - Controlador de disco (passivo)

Potência do sinal no barramento: os sinais binários no barramento são "fracos", por isso 3 conceitos estão atrelados a essa questão:

Controlador de barramento: conecta os dispositivos ativos ao barramento. Funciona como um amplificador de sinal.

Receptor de barramento: conecta os dispositivos passivos ao barramento.

Transceptor de barramento: conecta os dispositivos que são ativos e passivos ao barramento.

Ativo	Passivo	Exemplo
CPU	Memória	Buscar instruções e dados
CPU	Disp. E/S	Iniciar transf. de dados
CPU	Co-processador	CPU passando instruções para o co- processador
E/S	Memória	DMA (acesso direto a memória)
Co-processador	CPU	Co-processador buscando operando na CPU

- Projeto de um barramento (principais tópicos):
- largura do barramento
- relógio do barramento
- - arbitragem do barramento
- operações do barramento

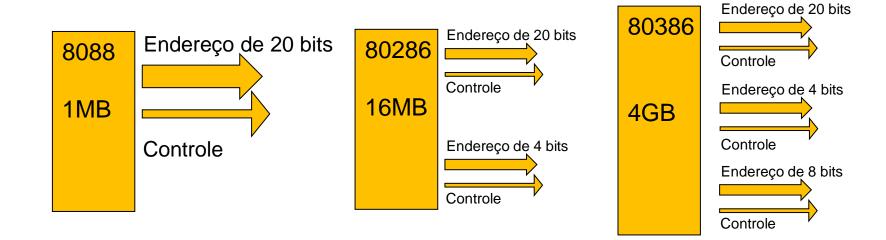
Largura do barramento!

Largura do barramento:

- -Linhas de endereço
- -Linhas de dados
- -Linhas de controle

Quanto mais linhas de endereço tiver um barramento, mais memória a CPU pode endereçar diretamente.

-"n" linhas de endereço, 2ⁿ endereços diferentes a serem localizados;



Largura do barramento:

Duas formas de aumentar a linha de dados:

- -Reduzir o ciclo de tempo de barramento (mais transferências por segundo)
- -Aumentar a largura do barramento (mais bits por transferência)

Problemas:

- acelerar o barramento é possível, mas difícil porque os sinais trafegam em linhas diferentes com velocidades ligeiramente diferentes, um problema conhecido como atraso diferencial do barramento.
- -Acelerar o barramento também causa incompatibilidade com dispositivos mais antigos.
- -Aumentar o número de linhas de dados aumenta o custo (mais cobre, maiores placas, etc) Mas é o mais usado entre as duas opções.

Relógio do barramento:

Barramento **síncrono**: suas linhas são comandadas por um oscilador de cristal.

- -Vantagem: projeto mais simples e custo menor.
- -Desvantagem: complicado de acompanhar a evolução da tecnologia. Caso uma nova memória com um tempo de acesso mais rápido seja lançada, esta terá que trabalhar na velocidade do barramento.

Barramento **assíncrono**: não possui um relógio mestre que comanda as atividades no barramento.

- -Vantagem: melhor adaptação ao avanço da tecnologia.
- -Desvantagem: projeto mais sofisticado e custo maior.

Arbitragem do barramento:

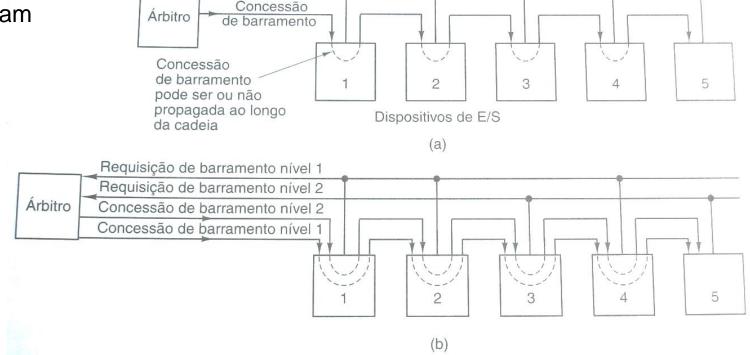
Mecanismo de arbitragem: usado para quando se tem dois ou mais dispositivos mestres informando o uso do barramento simultaneamente.

Arbitragem centralizada:

- (a) Dispositivo mais próximo sempre vence a disputa.
- (b) Divisão dos dispositivos em prioridades. Dispositivos
- 1, 2 e 4 usam prioridade 1 e dispositivo 3 e 5 usam prioridade 2.

Arbitragem descentralizada:

- cada dispositivo tem uma linha direta com o árbitro que julga a prioridade da requisição e libera o barramento para o dispositivo específico.



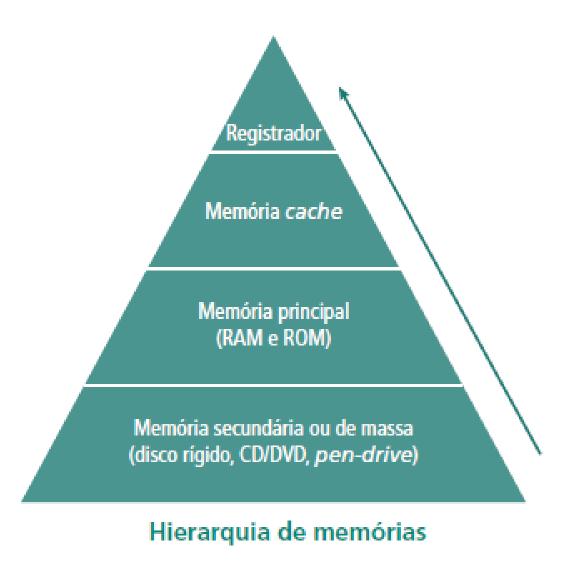
Requisição de barramento

Operações de barramento:

Normalmente uma só palavra é lida por vez, caracterizando uma operação no barramento.

Quando usado *caching* é preciso buscar uma linha inteira de cache (por exemplo, 16 palavras de 32 bits. Quando uma leitura de bloco é iniciada, o ativo informa ao passivo quantas palavras serão transferidas e o passivo faz a entrega, caracterizando outro tipo de operação no barramento.

Hierarquia de Memória



Alto custo
Alta velocidade
Baixa capacidade
de armazenamento

Baixo custo
Baixa velocidade
Alta capacidade
de armazenamento

Hierarquia de Memória

A memória cache é geralmente controlada por hardware

A memória principal (RAM) e a secundária é que o usuário tem acesso.

O sistema operacional através de um mecanismo de Memória Virtual (Segmentação e/ou Paginação) cria a "ilusão" ao usuário que a memória total é do tamanho da memória principal + memória secundária.

A técnica de memória virtual realiza transferência de blocos de informação entre a memória primária e secundária automaticamente sem a intervenção do usuário comum.

Registradores

São dispositivos (elementos computacionais) capazes de receber dados, mantê-los armazenados por uma curto período de tempo e transferi-los para outro dispositivo.

· São, portanto, elementos de armazenamento temporário.

Os registradores fazem parte da CPU.

São extremamente rápidos e armazenam grupos reduzidos de bits.

A memória é a parte do computador em que os programas e os dados são armazenados.

A memória principal (MP) armazena programas em execução e os dados utilizados por eles.

Sem uma memória na qual processadores possam ler ou escrever informações, o conceito de computador digital com programa armazenado não pode ser implementado.

 A CPU processa instruções que são obtidas na MP e os resultados são retornados à MP.

A unidade básica de memória é o bit (binary digit)

- Abstração de valores 0 ou 1
- Fisicamente é mais fácil distinguir entre dois valores distintos do que de mais valores. Tensão, corrente, ...

A memória é formada por um conjunto de células (ou posições), cada uma das quais podendo guardar uma informação.

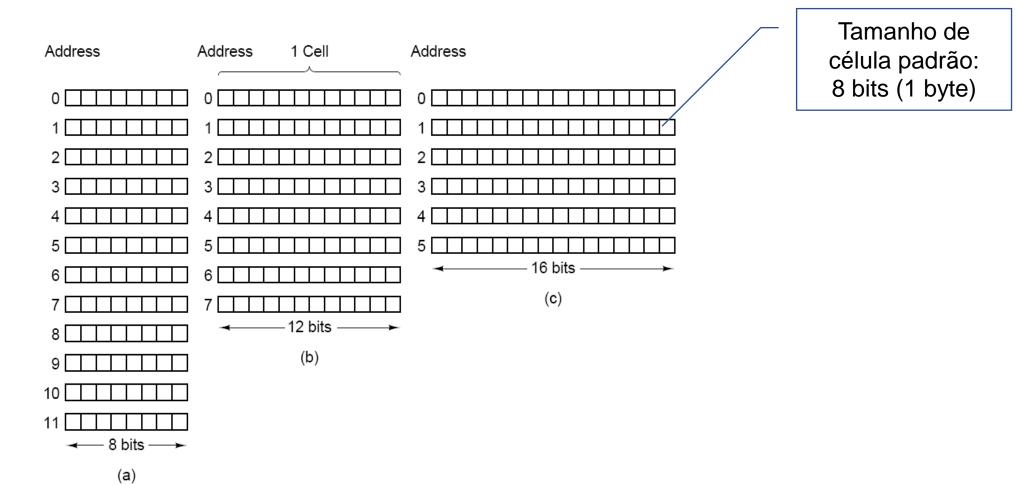
- Todas as células de uma dada memória têm o mesmo número de bits
- Os números que identificam (referenciam) a posição da célula na memória são chamados de Endereços.
- A célula é a menor unidade endereçável da memória

Endereços são indexadores pelos quais os programas podem referenciar dados na memória.

A memória é formada por um conjunto de células

Todas as células de uma memória têm o mesmo nº de bits

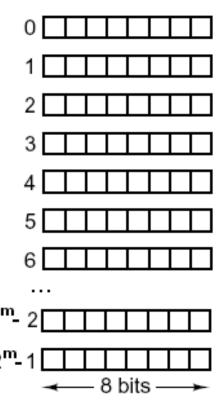
Cada célula tem um endereço



Se a memória tiver n células, elas terão endereços de 0 a (n - 1).

Se uma célula tiver k bits, ela poderá armazenar qualquer uma das 2^k combinações possíveis para os bits.

Se um determinado endereço tem m bits, o número máximo de células endereçáveis é de 2^m.



Os computadores modernos agrupam as células (ou bytes) em Palavras (word)

Ex: uma palavra de 32 bits tem 4 bytes (ou 4 células)

Nesses computadores, a Palavra é a parte mínima de dados que podem ser transferidos de/para a memória (MP)

A informação na palavra pode ser um dado ou uma instrução.

"Processadores de 32 bits" têm palavras de 32 bits

Registradores de 32 bits

nº de bits do barramento de endereços em geral (mas não obrigatoriamente) é igual ao nº de bits dos registradores (ex: Memory Address Register - MAR)

Instruções de 32 bits

Cada instrução deve tratar palavras de 32 bits

movimentar, somar ou subtrair dados armazenados em registradores de 32 bits

bytes de uma palavra podem ser ordenados na memória:

da esquerda para a direita
ou
da direita para a esquerda

————

Big Endian: bytes mais significativos primeiro (Mac)

Little Endian: bytes menos significativos primeiro (Intel)

00000000 00000000 00000100 00000001

	Big-Endian	Little-Endian
Address	representation	representation of
	of 1025	1025
00	00000000	00000001
01	00000000	00000100
02	00000100	00000000
03	00000001	00000000

O endereço identifica a palavra na memória, e é usado tanto para armazenar (store) como pará carregar (load).

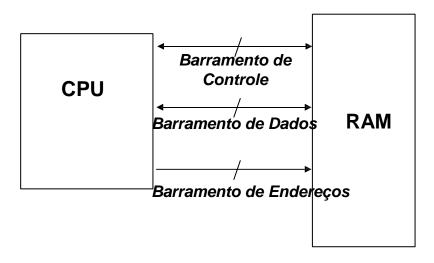
Em geral, registradores são usados para endereçar a memória tamanho do barramento de endereços = tamanho da palavra

No projeto de cada computador é definida a organização da memória (tamanho do barramento de endereços, tamanho da palavra e da célula de memória, etc.)

- Exemplos de memória:
 - 64K (2¹⁶) X 8 bits (65 536 posições de 8 bits = 64 kilobytes)
 - 4 Giga (2³²) X 8 bits (4 294 967 296 posições de 8 bits)
 - 1 Mega (2²⁰) X 16 bits (1 048 576 posições de 16 bits)

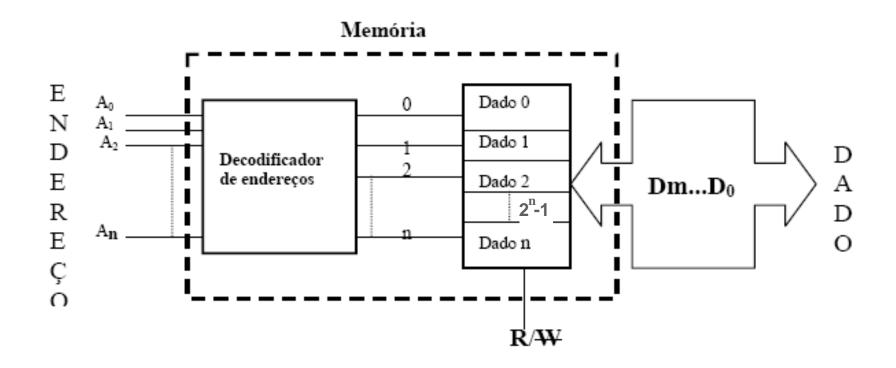
Leitura: Quando um valor é recuperado da memória, o conteúdo da palavra não é alterado. Apenas uma cópia será enviada pela memória.

Escrita: A gravação de um novo conteúdo em uma palavra se dá com a destruição do conteúdo anterior.



Organização da memória

Para simplificar: 1 palavra = 1 célula



- Uma memória com n linhas de endereços e m linhas de dados.
- As linhas de endereços (n bits) permitem endereçar 2ⁿ diferentes células de memória.
- O tamanho do dado, contido em cada célula de memória é de m bits.
- Cada código de endereço de n bits seleciona, através do decodificador de endereços, apenas uma única célula na memória.
- O sentido do fluxo de dados dependerá do sinal de controle R/W.
 - Se R/W = 1, a operação é de READ, ou seja, leitura na memória e o dado sairá da memória
 - Se R/W = 0, a operação será de escrita (WRITE) e o dado entrará na memória.

Tipos de Memória

Memória Volátil: É aquela cuja informação original é perdida se a energia for desligada.

Memória Não-volátil: É aquela que retém o padrão de bits original mesmo que a energia seja desligada.

Memória Endereçada Sequencialmente: Para se obter a informação de um endereço, é necessário percorrer os endereços anteriores.

Ex.: Fita Magnética

Memória de Acesso Randômico (RAM – Random Access Memory):

- A célula pode ser acessada sem ter que percorrer os endereços anteriores.
- O tempo de acesso é praticamente o mesmo para todas as células
- Volátil

Códigos com Correção/Detecção de Erros

Os dados armazenados na memória dos computadores podem ser "acidentalmente" alterados.

· Oscilações de tensão de alimentação, etc.

Para a prevenção desses erros, algumas memórias armazenam informações extras, usando códigos e mecanismos que permitam a detecção e/ou correção de erros.

Códigos com Correção/Detecção de Erros

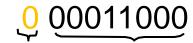
EXEMPLO

Bit de Paridade: bit configurado em 0 ou 1 para assegurar que o número total de bits 1 no campo de dados é par (ou ímpar).

Paridade refere-se ao número de bits '1' no número binário.

Ex:

bit de paridade dado



bit de paridade dado

Memória Cache

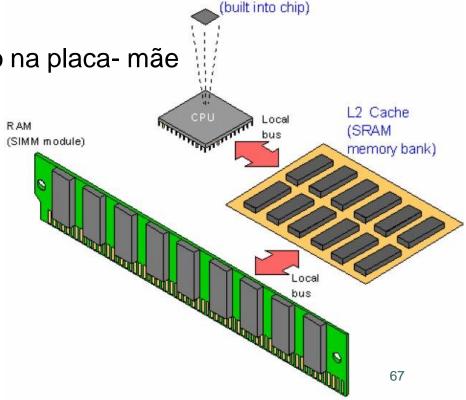
Quando um processador precisar de uma palavra de memória, ele primeiro busca essa palavra na cache.

Somente no caso de ela não estar armazenada na cache é que a busca se dará na memória principal.

Se uma parte substancial dos acessos for satisfeita pela *cache*, o **tempo médio** de acesso a uma palavra em memória será pequeno, próximo ao tempo de acesso à *cache*.

Em alguns computadores podem existir diversos níveis de cache

Ex: nível 1 é implementado dentro do chip ; nível 2 implementado na placa- mãe



L1 Cache

BOOT

Step 1.

A fonte envia um sinal a placa-mãe

Step 2.

Processador acessa o BIOS (Basic Input/Output Sistem) para inicializar o computador

Step 3.

BIOS roda o teste, chamado de POST, para testar componentes como mouse, teclado e placas conectadas

power-on self-test (POST)



Step 4.

Os resultados do POST são comparados com os dados no chip CMOS

Step 5.

BIOS verifica o sistema de arquivos em um drive USB, CD-ROM e então no HD

Step 6.

Kernel do sistema operacional é carregado para a memória RAM

Step 7.

O sistema operacional carrega as configurações informadas e mostra o desktop (área de trabalho) na tela